Dal record al dato. Linked data e ricerca dell'informazione nell'OPAC

Antonella Iacono

1 Linked data: la "struttura che connette" il web

> «Quale struttura connette il granchio con l'aragosta, l'orchidea con la primula e tutti e quattro con me? E me con tutti voi? E tutti e sei noi con l'ameba da una parte e con lo schizofrenico dall'altra?» (Bateson).

La citazione di Bateson in esergo si riferisce alla necessità, per chi fa ricerca, di allargare gli orizzonti del proprio pensare mettendo in relazione questioni o "cose" apparentemente lontane per ottenere una visione nuova e una comprensione più profonda della realtà.

La sua teoria "ecologica", che lo portava a cercare incessantemente nuove "strutture di collegamento", è un buon esempio per richiamare la necessità di superare la parcellizzazione del sapere contemporaneo e di giungere a nuove forme di apprendimento in grado di cogliere la complessità delle relazioni tra le cose del mondo che ci circonda.

Oggi nel web si assiste a un profondo rinnovamento delle strutture logiche di connessione dell'informazione che si sono evolute

JLIS.it. Vol. 5, n. 1 (Gennaio/January 2014).
DOI: 10.4403/jlis.it-9095

dal collegamento ipertestuale di documenti scarsamente strutturati fino a quello logico delle unità minime di significato, i dati. Il web ha trovato nei linked data una nuova "struttura che connette" l'informazione presente in rete e che rende possibili collegamenti inediti tra i diversi domini del sapere gettando le basi per nuove forme di apprendimento e per una conoscenza interdisciplinare e contestualizzata.

Il termine linked data, nella definizione classica fornita da Tim Berners Lee, fondatore del Web e principale esponente del movimento del Web semantico, si riferisce ad una serie di "buone pratiche" per pubblicare e collegare dati strutturati sul web e formare così la base per il Web semantico.[1]

Linked data significa però, soprattutto, un mutamento di paradigma: vuol dire creare dati che siano "del web" e non solo "sul web" (Coyle) ovvero abbracciare un nuovo modo di creare informazione strutturata che consente ad ogni tipo di informazione di essere strettamente collegata e interoperabile con ogni altra informazione presente nel web.

La prospettiva del Web Semantico coinvolge direttamente lo sviluppo di sistemi di *information* retrieval come i cataloghi elettronici e le modalità di ricerca possibili, consentendo il superamento dell'*information retrieval* tradizionale, facilitando il ritrovamento dei contenuti dei documenti stessi e favorendo, al contempo, l'integrazione con altri documenti presenti nel web con la creazione di rete diffusa di conoscenze. In questo contributo, che si ricollega a quello

[1] Com'è noto con i linked data è possibile pubblicare i dati in una modalità leggibile e interpretabile da una macchina, il cui significato è definito da una stringa di parole e marcatori (tripla) per costituire un reticolo di dati collegati appartenenti ad un dominio. Questi sono collegabili ai *dataset* relativi agli altri domini presenti nel web costruendo in tal modo una rete di dati globale, i cui contenuti possono essere scambiati e interpretati dalle macchine, ovvero la base per il Web Semantico (Heath e Bizer; Guerrini e Possemato)

già pubblicato su questa rivista (Iacono), si esplorano le potenzialità e i benefici dell'applicazione dei linked data alla strutturazione logica dei dati bibliografici nel catalogo, o meglio in quello strumento che alla luce dei più recenti sviluppi (Marchitelli e Frigimelica) può essere più propriamente definito un dispositivo per l'accesso e la mediazione dei contenuti della biblioteca, con un'attenzione alle funzionalità di ricerca.

Circoscrivendo la presente osservazione alle potenzialità per la ricerca, il principale vantaggio che può derivare dall'applicazione di questa nuova struttura logica al catalogo risiede infatti nella capacità di contestualizzare l'informazione creando collegamenti tra le conoscenze durante il processo informativo. Il catalogo può diventare lo strumento con cui le biblioteche contemporanee possono colmare le lacune di un sapere oggi sempre più frammentato e agevolare i processi di trasformazione delle informazioni in conoscenza collegando le conoscenze prodotte dalle biblioteche agli altri dati presenti nel Web (Di Domenico).

Con le iniziative intraprese dalle biblioteche nel campo dei linked data ci si avvia verso una nuova importante tappa evolutiva che vedrà verificarsi una radicale trasformazione del catalogo e dei dati bibliografici, del rapporto tra utente e catalogo e tra catalogo e universo globale dell'informazione. Affinché i record bibliografici siano in grado di giocare un rinnovato ruolo nel web è però indispensabile che i dati delle biblioteche non vengano solamente esposti nel nuovo formato Resource Description Framework (RDF), bensì che le basi dati catalografiche adottino RDF come struttura del catalogo (Malmsten), compiendo il fondamentale passaggio dal record al dato.

2 Dal record al dato. Il nuovo "record bibliografico collegato" nella gestione del processo informativo

Il *record* bibliografico, principale strumento per rappresentare l'informazione bibliografica nel catalogo, è un insieme di metadati che rappresentano le entità e gli attributi che caratterizzano le dimensioni semiotiche e semantiche di un "oggetto bibliografico" (una monografia, un documento sonoro, etc.) e le relazioni con le altre entità collegate.

Per molto tempo la produzione di record bibliografici si è basata su una struttura piatta e lineare del record, fornita dai formati Machine Readable Cataloguing (MARC), oggi ritenuta obsoleta (Bianchini; Kroeger) e non più adatta a rappresentare compiutamente sia le diverse entità (opera, espressione, manifestazione, item, soggetti, autori, enti, famiglie), sia i legami tra le entità previsti dai modelli IFLA. Più volte l'attuale struttura del record bibliografico è stata vista come il principale ostacolo all'espressione delle entità rappresentate nel catalogo (Yee); un record bibliografico che manca di una granularità adeguata rende difficoltoso lo svolgimento delle basilari funzioni del catalogo (*funzioni-utente*).[2] I linked data offrono la possibilità di operare una profonda ristrutturazione del record bibliografico che si presenta con una nuova struttura granulare e

[2]Lo sviluppo del MARC è stato di fatto abbandonato; la Bibliographic Framework Transition Initiative (http://www.loc.gov/marc/transition), avviata nell'ottobre 2011 dalla Library of Congress, ha come principale attività lo sviluppo di nuove modalità di creazione e condivisione dei record bibliografici e la ricerca di una soluzione per la sostituzione del formato MARC per lo scambio dei dati bibliografici. Il progetto di un nuovo Bibliographic framework (BIBFRAME, http://www.loc.gov/bibframe e http://bibframe.org) come nuova modellazione dei dati bibliografici si centra sull'ambiente Web e adotta il data model RDF, i principi e le specifiche dei linked data (Marcum; Kroeger).

atomizzata; il nuovo concetto di record bibliografico si esprime nel "record collegato", costituito virtualmente dalla sintesi di varie triple che combinano dati creati dalle biblioteche e da altri enti.

Attraverso questa rinnovata struttura logica è possibile generare un record formato dalle entità già previste dai modelli concettuali FR dell'IFLA[3] e dotare il record bibliografico di una rete di collegamenti che consentirà all'utente di svolgere efficacemente le funzioni di base del processo informativo.[4] Possiamo, dunque, immaginare il nuovo "record collegato" come prosecuzione dell'idea di "super-record" o "meta record" (Fattahi, *The relevance of cataloguing principles to the online environment*; *From Information to Knowledge*), recentemente riproposta come una struttura che possa essere agevolmente navigata in modo da fornire all'utente «un ambiente di apprendimento per mezzo del quale il *record* catalografico facilita l'organizzazione e la rappresentazione della conoscenza nel catalogo, come pure la capacità di acquisire, elaborare e interiorizzare la conoscenza da parte di chi utilizza il catalogo». (*From Information to Knowledge*, p. 37).

Il nuovo "record collegato", che si viene a creare quando i dati del catalogo vengono strutturati come linked data, è formato da un collegamento di triple in grado di connettere dati provenienti da diverse fonti sia bibliografiche sia di altra natura, sfruttando un reticolo illimitato di collegamenti tra i dati, il *Web of data*. L'elemento che consente di realizzare questo collegamento è il concetto di Uniform Resource Identifier (URI) ovvero un identificatore unico e persistente associato al dato. Nel nuovo "record collegato" ogni par-

[3] Anche lo sviluppo dei modelli FR si sta orientando, non senza difficoltà, verso una rappresentazione ontologica adatta al web semantico; va ricordato in proposito lo sforzo verso la trasformazione di Functional Requirements for Bibliographic Records (FRBR) nella struttura di FRBR-object oriented (FRBRoo), attraverso CIDOC CRM (Le Boeuf, «Foreword»; «A Strange Model Named FRBRoo»).

[4] Per un approfondimento sulle fasi del processo informativo e sulle funzioni che l'utente compie durante l'interazione con l'OPAC si veda Iacono.

te della tripla viene identificata attraverso un URI; avremo dunque URI che identificano le opere, le espressioni, gli autori, i soggetti e le classi, ma anche le relazioni che intercorrono tra le entità saranno a loro volta provviste di URI in modo tale da poter essere riutilizzate. Si crea così un reticolo di dati in cui tutte le entità e le relazioni previste dai modelli FRBR possono essere espresse e nel quale le entità del catalogo possono essere relazionate alle altre entità presenti nel web. Ciò è reso possibile dal recente sviluppo di set di elementi, di vocabolari e ontologie predisposti dalle biblioteche (es. Dublin Core, RDA) [5] o da altre comunità che operano nel Web (Foaf, SKOS). Il collegamento tra i dati di diverso tipo viene creato automaticamente tramite la connessione di diversi dataset con uno sforzo che non ricade unicamente sulle biblioteche, ma viene condiviso tra le biblioteche e le altre comunità del Web Semantico.

3 La riduzione delle barriere cognitive della ricerca e il recupero dell'informazione nel contesto.

Vi sono almeno tre aspetti dell'interazione utente-catalogo che possono beneficiare dell'applicazione dei linked data alla struttura catalografica: la riduzione delle barriere cognitive che intervengono nel processo di ricerca, l'accesso ai dati bibliografici e il miglioramento della ricerca semantica. Nell'interazione che si compie nel catalogo l'utente incontra numerosi ostacoli sia al momento dell'attivazione

[5] Le iniziative delle biblioteche nell'ambito del Web Semantico si sono concretizzate nella creazione di dataset che hanno che messo a disposizione i dati di cataloghi e bibliografie nazionali in RDF (British National Bibliography, Data.bnf. fr, Biblioteca Nacional de España, Deutsche Nationalbibliografie (DNB) e HBZ Union Catalog, Libris), vocabolari (LCSH, Dewey, LCC, Rameau, Dewey, VIAF) e schemi di metadati (RDA, Dublin Core, FRBR, ISBD etc.).

della ricerca sia lungo il processo informativo (Kulthau; Wilson; Godbold).[6] Tra le barriere maggiormente significative vi sono quelle cognitive (Borgman), che possono consistere:

- in una scarsa comprensione dell'oggetto della ricerca, il focus, presente qualora l'utente si trovi in una fase iniziale della sua ricerca;
- in una difficoltà di espressione dell'oggetto di ricerca, derivante dalla scarsa conoscenza dell'area disciplinare che si sta indagando (domain knowledge);
- nella scarsa conoscenza di un sistema e dei suoi meccanismi di recupero dell'informazione (barriere tecnologiche semantiche e concettuali);
- nella scarsa conoscenza dell'universo bibliografico rappresentato e delle modalità di rappresentazioni dei testi;
- nelle barriere linguistiche.

A queste si possono aggiungere anche quelle barriere provenienti dai modelli mentali che gli utenti hanno costruito nell'utilizzo di altri strumenti di ricerca (Yu).

L'adozione dei linked data come struttura del record bibliografico riduce l'incidenza di questa tipologia di barriere creando le basi per un processo di apprendimento che è fondamentale per il successo di una ricerca nel catalogo. Diviene centrale il concetto di "contesto", concetto che è profondamente connaturato nella definizione di conoscenza, in quanto la conoscenza stessa viene creata e

[6] Tra le barriere rientrano quelle fisiologiche, cognitive, emozionali, quelle che fanno riferimenti a condizionamenti sociali ed economici (accesso e costo dell'informazione), quelle ambientali e situazionali (tempo a disposizione, fattori geografici, linguistici) e culturali ai quali auspica possa essere riservato un più ampio spazio nella progettazione dei sistemi di ricerca.

appunto collocata in un contesto. Una progettazione efficace del catalogo dovrebbe creare un contesto nel quale l'utente può collocare ogni singola informazione per mezzo della relazione tra le entità costitutive. Il principale ruolo dei linked data nell'innovazione dei cataloghi elettronici risiede quindi nella possibilità che il "web dei dati" venga utilizzato come "contesto esteso" per l'esplorazione dei contenuti che si affianca a quello rappresentato nel catalogo (ossia l'universo delle entità e dei loro collegamenti), per espandere e migliorare il processo di ricerca dell'informazione. I linked data consentono di appoggiare il catalogo ad una infrastruttura cognitiva costituita dalle conoscenze presenti nel "web dei dati", che collega le conoscenze formatesi in contesti diversi.[7]

Tale contesto allargato è in grado di migliorare le seguenti funzioni dell'utente:

Navigazione e browsing nel catalogo. L'utente che si trova al momento dell'interazione con l'OPAC nelle prime fasi del processo informativo è naturalmente predisposto a una ricerca esplorativa preferendo modalità attive ma indirette di esplorazione. La tecnologia dei linked data offre varie opportunità, prima fra tutte quella di migliorare i percorsi di navigazione e di approccio al catalogo. Attraverso i linked data è infatti possibile rappresentare le entità catalografiche e consentire una navigazione gerarchica che dal raggruppamento delle opere si svolga fino all'item. In tal modo si compiono le "funzioni bibliografiche" che l'utente svolge nel catalogo e si realizza una navigazione significativa per l'utente e, al contempo, un processo di apprendimento. La navigazione attraverso il colle-

[7] Lo schema di un reticolo bibliografico espresso con i linked data nell'OPAC Libris della Biblioteca nazionale svedese è visibile all'URL: http://librisbloggen.kb.se/2008/12/03/libris-available-as-linked-data Un data model è quello della British Library visibile in http://www.bl.uk/bibliographic/pdfs/bldatamodelbook.pdf.

gamento ad altri *dataset* (es. DBPedia) può estendersi al web, inglobando dinamicamente informazioni provenienti da fonti esterne.

Interrogazione tramite formulazione di una query. Nelle fase intermedia del processo di ricerca, in cui l'utente è in grado di formulare interrogazioni al sistema, è invece estremamente importante offrire un aiuto nel formulare *query*. Sebbene gli OPAC oggi ricorrano all'autocompletamento, sovente questa tecnica non si appoggia a vocabolari controllati, rivelandosi inefficace. I linked data, che cominciano ad avere alcune prime applicazioni nello sviluppo dei cataloghi elettronici (Malmsten).[8] Vi sono anche prime esperienze di integrazione in discovery tools open source come VuFind (Hatop), consentono di usufruire di ontologie già create in maniera dinamica e automatica all'interno del Web Semantico e dunque utilizzabili a vantaggio della ricerca nei cataloghi per aiutare l'utente a formulare interrogazioni. Tramite i legami tra diversi *dataset*, si offre all'OPAC una struttura sindetica potenzialmente illimitata che può servire da vocabolario d'accesso al catalogo. Ad esempio, gli accessi semantici di un catalogo (come le stringhe LCSH o le notazioni delle classi) vengono automaticamente mappati ai concetti correlati e diventano recuperabili non solo a partire dai termini che compongono il soggetto, ma anche a partire da termini semanticamente simili, perché collegati allo stesso concetto. La rete dei dati presente nel Web dunque potrebbe utilmente essere sfruttata dalle biblioteche per sviluppare interfacce che aiutino l'utente a scegliere il soggetto più appropriato per la propria ricerca. La tecnologia dei linked data consente anche di abbattere le barriere linguistiche all'utilizzo di un catalogo, in quanto tutte le forme varianti

[8] Un esempio è l'OPAC Libris http://libris.kb.se.

di un nome, di un titolo o di un soggetto possono essere collegate alla forma accettata secondo il meccanismo già noto dell'*authority file* con un vantaggio in più per la ricerca semantica: la possibilità di esprimere un concetto in una lingua diversa da quella per cui il catalogo è progettato; i termini dell'interrogazione sebbene in una lingua diversa da quella espressa in un catalogo possono venire ugualmente collegati automaticamente tramite i linked data con un indubbio vantaggio per chi ricerca. Pensiamo alla possibilità di collegare i dati degli archivi di autorità: collegando il dato alla forma in VIAF, [9] l'utente potrebbe trovare una descrizione catalografica partendo dalla forma desiderata tramite il collegamento a tutte le forme varianti. Allo stesso modo ogni dato può essere collegato o provenire da archivi di diverso tipo: statistici, di anagrafi pubbliche, di enti territoriali o governativi, etc.[10] Un ulteriore beneficio è quello di poter interrogare il catalogo servendosi del linguaggio naturale per essere indirizzati ai termini più adatti per la ricerca nel catalogo o lanciare una ricerca esplorativa a partire dai termini suggeriti dal sistema.

Valutazione dell'informazione ricevuta. La valutazione dei risultati di un'interrogazione o dei record recuperati nel corso della ricerca nel catalogo consiste, per l'utente che si trova nelle prime fasi del processo di ricerca, essenzialmente nel valutare la rispondenza all'argomento ricercato (rilevanza topica), mentre per chi si trova già nelle fasi successive anche alla corrispondenza soggettiva, ovvero un ultima analisi alla pertinenza e

[9] http://www.viaf.org.

[10] Tra i progetti è da citare l'arrichimento di degli articoli di Wikipedia con i record d'autorità VIAF da parte dei wikipediani di OCLC e British library (Klein e Alex), esperimento che viene replicato nel nostro paese dai bibliotecari wikipediani con i gli authority records SBN; l'operazione sarebbe ancor più interessante se ampliata alle voci d'autorità per i soggetti.

all'utilità. L'opportunità offerta dalla tecnologia dei linked data tramite il record bibliografico "collegato" risiede nella possibilità di ampliare il contesto utile all'utente per effettuare la valutazione della rilevanza e della pertinenza, tramite il supporto delle migliaia di risorse presenti nel Web Semantico. La quantità di dati recuperabile dinamicamente attraverso un record collegato al *Web of data* è non solo potenzialmente illimitata, ma arricchisce il risultato di informazioni supplementari sempre aggiornate, migliorando la valutazione: ad esempio la possibilità di visualizzare un raggruppamento di tutti gli argomenti trattati da un autore nelle sue opere, di visualizzare la fortuna editoriale di un autore in un dato periodo di tempo o l'autorevolezza, sono funzionalità tutto sommato non particolarmente complesse, ma di grande aiuto all'utente, che possono essere implementate per mezzo del collegamento del record ai dati presenti nel Web.[11]

Accesso alle risorse bibliografiche. La novità risiede nella possibilità di accedere al catalogo da fonti esterne; una volta che i dati bibliografici entrano a far parte del Web of data essi saranno oggetto del recupero dell'informazione da qualsiasi altro dato collegato al *Web Semantico*.

Visualizzazioni avanzate. Una delle novità introdotte dagli OPAC di nuova generazione è certamente la componente visuale del recupero dell'informazione, manifestatasi non solo nella

[11] In questa direzione un'iniziativa interessante è Social Networks and Archival Context project (SNAC), un sistema per la ricerca storica a partire dai record archivistici di enti, persone e famiglie che utilizza i linked data per ricostruire il contesto storico e sociale (persone, enti date, luoghi, eventi, soggetti correlati, cronologie) e del quale un prototipo è visibile in http://socialarchive.iath.virginia.edu/index.html. Dei collegamenti al *Web semantico* si serve oggi anche Google per presentare accanto alla SERP i risultati provenienti estratti da Wikipedia.

comparsa delle copertine dei volumi quale arricchimento del record, ma soprattutto nella predisposizione di nuove modalità di browsing visuale delle collezioni e nella costruzione di canali di esplorazione delle risorse. Non c'è dubbio che oggi molte applicazioni del web si distinguano anche per la loro veste accattivante che si esprime, sovente, nella creazione di mappe visive. Il record bibliografico 'collegato', che la logica dei linked data consente di costruire, getta le basi per la costruzione di «ambienti arricchiti di conoscenza» (Fattahi, *From Information to Knowledge*, p.36) nei quali sarà possibile per l'utente attivare nuovi percorsi di esplorazione del catalogo a partire dalle entità oggetto di interesse. Non dobbiamo infatti dimenticare che le nuove possibilità di implementazione di FRBR ai cataloghi renderanno possibile in futuro organizzare i risultati delle ricerche tramite raggruppamenti logici previsti dal modello. Il "record collegato" compie un ulteriore passo in avanti: attraverso il collegamento dei dataset delle biblioteche con quelli presenti nel *Web dei dati* si può dare luogo alla realizzazione di rappresentazioni ontologiche dei risultati delle ricerche che gli utenti mettono in atto, capaci di aprire orizzonti inaspettati per la ricerca semantica.

4 I Linked data e i nuovi orizzonti della ricerca semantica

Il *data model* fornito da RDF, che è alla base dell'architettura dei linked data, consente di strutturare l'informazione bibliografica in modo da potenziare le funzionalità della ricerca semantica. Ciò è possibile grazie al fatto che nel Web Semantico tutti gli URI che si riferiscono allo stesso concetto sono correlati creando una vasta rete

semantica di relazioni che può essere automaticamente collegata al catalogo. Come è noto, i cataloghi soffrono della contingenza e della difficoltà per l'utente dovuta all'utilizzo dei dispositivi indicali: essi rappresentano una necessità, ma pongono all'utente il problema di misurarsi con il "linguaggio bibliografico" del catalogo (Serrai, *Indici, logica e linguaggio: problemi di catalogazione semantica*; *Del catalogo alfabetico per soggetti: semantica del rapporto indicale*; Ventura). Per questo motivo gli accessi semantici sono oggi sempre meno utilizzati dagli utenti che preferiscono immettere parole chiave ed effettuare una ricerca *Google like* su tutti i campi ottenendo sovente risultati imprecisi.

Gli indici che le biblioteche utilizzano ai fini dell'organizzazione delle registrazioni catalografiche e del loro recupero riflettono tuttavia il modo in cui le biblioteche organizzano il sapere. Per tradizione esse hanno seguito strategie di indicizzazione *content oriented* ovvero tese ad individuare con precisione nei termini indicali il contenuto di un documento in base all'utenza di riferimento, sacrificando così le potenzialità di utilizzazione dei libri stessi in quanto le molteplici dimensioni e interpretazioni che possono provenire da studiosi di formazione e provenienza scientifica diversa vengono ridotte alla creazione di un'unica stringa coestesa (Biagetti).[12]

[12]Biagetti analizza il ruolo degli indici semantici nel contributo Indici semantici e ricerca scientifica che, ad una rilettura alla luce delle nuove tecnologie del web semantico, offre interessanti spunti di riflessione. La studiosa rileva come rispetto agli aspetti sintattici del processo di indicizzazione sia stata tradizionalmente riservata una minore attenzione all'analisi dei contenuti semantici e alla scelta delle strategie di indicizzazione da adottare. Ciò è dovuto – afferma la studiosa – al prevalere di metodologie che si sono richiamate al procedimento della coestensione ovvero alla necessità di condensare in una sola stringa i molteplici contenuti semantici di un libro. A tale strategia orientata al contenuto *content oriented* si contrappone, specie nell'indicizzazione di documenti che ricadono nelle Scienze umane, la possibilità di studiare strategie need-oriented, ovvero indirizzate alle necessità e alle possibili richieste di ricercatori formatisi in contesti scientifici diversi (, p.16 «Indici semantici e ricerca scientifica»).

Oggi i Linked data ci offrono l'opportunità di collegare le sistemazioni concettuali messe in opera dalle biblioteche alle altre modalità di organizzare e rappresentare le conoscenze create da altre comunità che operano nel web:

> «An essential process is the joining together of subcultures when a wider common language is needed. Often two groups independently develop very similar concepts, and describing the relation between them brings great benefits... The Semantic Web, in naming every concept simply by a URI, lets anyone express new concepts that they invent with minimal effort. Its unifying logical language will enable these concepts to be progressively linked into a universal Web» (Berners-Lee, Hendle e Lassila).

L'affermazione di Berners Lee, Hendler e Lassilla esprime l'enorme potenzialità dell'applicazione del Web Semantico ai cataloghi: il fatto di essere una logica unificante che consente di collegare i diversi linguaggi che nel web vengono utilizzati per denominare cose, eventi, luoghi, concetti, persone. Le principali potenzialità applicative che i linked data offrono alla ricerca semantica sono ravvisabili in almeno due principali punti:

- mappatura di concetti presenti in schemi e sistemi di organizzazione delle conoscenze diversi ai fini dell'interrogazione, della visualizzazione e della navigazione nel catalogo;
- sviluppo di modalità multidimensionali per la ricerca.

Il Web semantico fornisce un'infrastruttura tecnologica che consente di collegare allo stesso concetto identificato da un URI i termini o le espressioni utilizzate da diverse comunità del Web ottenendo una rete sindetica già pronta che le biblioteche possono utilizzare per i

propri scopi. La scomposizione dei soggetti e delle classi nei concetti che li compongono potrà unire diversi sistemi di organizzazione delle conoscenze (KOS) creati dalle biblioteche e sistemi creati da altre comunità del web semantico, offrendo modalità avanzate di recupero dell'informazione che si possono avvalere di una rete di rimandi tra termini già pronta all'uso. Sarà dunque possibile raggiungere un soggetto o una classe a partire da termini provenienti da KOS di diverso tipo e collegare tra loro termini affini e concetti simili ampliando il raggio della ricerca semantica con indubbi vantaggi per l'utente. Il *data model* fornito da RDF è molto simile a quello delineato da FRSAD, lo schema concettuale di IFLA per modellare la circalità delle opere (*aboutness*) che adotta un modello fondato sulle entità *thema* (concetto) e *nomen* (etichetta).[13] FRSAD trova adeguate corrispondenze nei due schemi di codifica per il Web semantico: Simple Knowledge Organization System (SKOS) e Web Ontology Language (OWL),[14] che forniscono modelli per esprimere le strutture di sistemi di organizzazione delle conoscenze.

In particolare, SKOS definisce classi e proprietà per esprimere la struttura di base e il contenuto di schemi concettuali come thesauri, schemi di classificazione, tassonomie, soggettari, folksonomies e altri vocabolari controllati. Questo standard adotta un modello basato sul "concetto" e sulla separazione tra concetti e etichette (ovvero i termini per esprimerli) già espressa in FRSAD; il suo vocabolario è modellato sui concetti che a loro volta sono rappresentati da etichette e che sono confrontabili con le entità definite in FRSAD ovvero *thema*, *nomen* e i loro attributi. Attraverso SKOS i concetti possono essere organizzati in gerarchie e reti di associazioni, aggregati in schemi

[13] In particolare si veda IFLA, FRSAD Final Report: Appendix C: FRSAD Model and Other Models, p. 48-50.

[14] Rispettivamente http://www.w3.org/2009/08/skos-reference/skos.html#ConceptScheme e http://www.w3.org/TR/#tr_OWL_Web_Ontology_Language

concettuali, ordinati in collezioni ordinate e mappati ai concetti presenti in altri schemi (Žumer e Zeng).

Il modello definito da FRSAD trova mappature appropriate nelle classi, negli attributi e nelle relazioni semantiche stabilite dallo standard del W3C SKOS: frsad:thema equivale a skos:Concept; frsad:nomen equivale a skos:label e altre proprietà come altre proprietà come skos:broader, skos:narrower, skos:related consentono di esprimere le relazioni gerarchiche e associative tra i concetti tipiche di un'organizzazione thesaurale. Nella costruzione di una tripla ogni concetto viene associato – tramite il predicato rdf:type – alla classe specifica skos:Concept ed identificato in modo univoco da un URI. Lo standard prevede la possibilità di esprimere un'etichetta in diverse lingue, di identificare i termini che all'interno dello schema concettuale sono accettati e usati come descrittori nel sistema d'indicizzazione, e infine, di descrivere anche gruppi di termini che sono di corredo al concetto e che non hanno un'identità semantica e che quindi non verranno identificati con un URI, ma solo con etichette (es. i sinonimi o i quasi sinonimi le varianti ortografiche, le abbreviazioni e gli acronimi).[15] Web Ontology Language (OWL) è un altro linguaggio per esprimere ontologie di concetti nel Web semantico con un significato formalmente definito. Esso permette di esplicitare le relazioni tra vari concetti (*thema*) previsti da FRSAD in maniera più completa rispetto a SKOS in quanto consente di esprimere ulteriori relazioni tra le classi come SubClassOf, EquivalentClasses, DisjointClasses, e DisjointUnion. I linguaggi SKOS e OWL consentono in combinazione di esprimere "concetti" ovvero accessi per

[15]Questo tipo di relazioni tra etichette possono essere risolte ricorrendo a un'apposita estensione di SKOS, SKOS-XL, che fornisce un supporto per descrivere entità lessicali (termini) attribuiti ai concetti. Lo scopo è reificare le etichette di skos:Concepts, trattandole come risorse RDF effettive. Ciò consente la possibilità per esse di essere ulteriormente annotate utilizzando ad esempio una proprietà "è una traduzione di". Lo schema è visibile in: xmlns:xl=http://www.w3.org/TR/skos-reference/skos-xl.html

soggetti e per classi e le loro relazioni, ma anche di mappare concetti presenti in diversi schemi.

L'assegnazione di un identificatore univoco ai concetti facilita l'interoperabilità tra KOS differenti, cioè la possibilità di mappare entità semantiche facenti parte di schemi concettuali diversi. Per stabilire questi collegamenti lo standard SKOS definisce alcune particolari proprietà che prevedono differenti livelli di equivalenza (skos:closeMatch; skos:exactMatch; skos:broaderMatch e skos:narrowerMatch; skos:relatedMatch). Di recente, i più importanti e diffusi soggettari e schemi di classificazione sono stati tradotti nella forma di Linked data, tra questi i vocabolari della Library of Congress (LCSH, LCC), la Classificazione Decimale Dewey, RAMEAU e il Thesaurus del Nuovo Soggettario consentendo questa tipologia di collegamenti. Sarà così possibile collegare schemi diversi di organizzazione delle conoscenze (KOS) creati dalle biblioteche come soggetti e classi[16] o schemi completamente diversi per tipologia e profondità come un soggettario e una folksonomia offrendo nell'OPAC il collegamento automatico dei record di autorità a varie altre forme presenti sul web che possono diventare, se collegate ai dati catalografici, punti d'accesso aggiuntivi al catalogo. Attraverso il collegamento e la mappatura di concetti presenti in differenti schemi la ricerca per soggetto può essere enormemente potenziata, sia nella fase di interrogazione, in quanto è ormai possibile utilizzare la rete di collegamenti del record per aiutare l'utente a formulare interrogazioni, sia in fase di navigazione in quanto è possibile produrre visualizzazioni avanzate che sfruttano il collegamento tra schemi

[16]Questo tipo di collegamento, che non viene ancora implementato negli OPAC, è un espediente che favorisce la navigazione e l'apprendimento che si 6può realizzare durante il processo informativo. L'utente può infatti partire da un soggetto per visualizzare tutte le classi che utilizzano quel soggetto e navigare nell'albero classificatorio correlato, o al contrario visualizzare tutti i soggetti che sono legati alle opere che fanno parte di una determinata classe

utilizzati sia nel settore bibliotecario che in altri settori scientifici, per espandere ulteriormente la ricerca. È inoltre possibile offrire all'utente l'opportunità di interrogare il catalogo secondo una logica sfumata: la possibilità di collegare concetti non identici ma simili, prevista nel Web Semantico, rende realizzabile il recupero di un concetto partendo da concetti ad esso correlati. In questo modo viene migliorata la ricerca semantica e la fruizione dei documenti stessi, che potranno essere più facilmente recuperati e utilizzati.

Queste nuove possibilità aprono prospettive inedite per la ricerca semantica sia in ambito generalista che in ambito scientifico: se le biblioteche, in base alla conoscenza della propria utenza di riferimento, hanno adottato per molto tempo strategie di indicizzazione semantica rivolte alla sommarizzazione estraendo dal documento i contenuti per ricavarne una circalità (*aboutness*) adatta ai lettori di riferimento, oggi invece, nei cataloghi elettronici la cui utenza è remota e non conoscibile a priori, si registra una mediazione minima da parte del catalogatore (Biagetti). Questa pratica – affermava giustamente la studiosa – contrasta con la visione pragmatista del soggetto espressa da Hjørland nella teoria della domain analysis (analisi di dominio) secondo la quale le proprietà che distinguono semanticamente un documento possono avere significati diversi all'interno di diversi ambiti disciplinari e che richiederebbe l'applicazione di modalità di indicizzazione *user oriented*.

La teoria del soggetto come «potenziale epistemologico dei documenti» elaborata dallo studioso danese (citato in Biagetti), afferma che le proprietà significative del soggetto non possono essere stabilite *a priori*, ma dipendono dal contesto nel quale vengono elaborate, ovvero dalla società stessa e che lo sviluppo epistemologico di un soggetto è frutto dell'avanzamento scientifico e sociale. In questo senso emergono gli elementi di novità e le nuove opportunità che la tecnologia dei linked data offre alla rielaborazione di una rinnovata

"teoria del soggetto":[17] il passaggio da una rete di "link" ad una rete di "significati" che si ottiene con la creazione del Web semantico, consente di allargare il raggio di analisi alle modalità di rappresentazione della conoscenza in diverse comunità scientifiche ed equivale alla possibilità per i nuovi sistemi di recupero dell'informazione di usufruire di quella struttura di significati che costituisce il portato epistemologico di un soggetto, creando le condizioni per proficui scambi disciplinari e per un recupero dell'informazione qualitativamente differente. Tornando alla ricerca nell'OPAC, particolarmente interessanti si rivelano i salti dimensionali che l'utente può compiere, che introducono nella ricerca semantica funzionalità decisamente innovative rese possibili dal collegamento di *data set* bibliografici a dataset di altra natura (ad esempio, quelli di persone, geografici, o tematici es. di film, di opere teatrali) o di altre istituzioni del patrimonio culturale come musei e archivi. Tramite il collegamento con i *data set* del patrimonio culturale sarà possibile per l'utente saltare da un soggetto ad un oggetto museale che lo rappresenta, (ad esempio un quadro, una scultura). Sarà, infine, possibile attivare collegamenti tra dati bibliografici alle "cose" reali descritte nel web; ad esempio un utente, attraverso il collegamento tra *data set* delle biblioteche e database geografici, potrà visualizzare in una mappa geografica i soggetti correlati o gli autori che pubblicano in un determinato paese o che hanno avuto più influenza in una certa area geografica. Un'ulteriore possibilità offerta dalle tecnologie del web semantico riguarda la progettazione delle interfacce dell'OPAC: la visualizzazione dei dati bibliografici potrà avvalersi delle tecnologie del Web semantico che consentono, ad esempio, la

[17] Così si esprimeva in proposito Biagetti (p. 24): «La Biblioteconomia e la Scienza dell'informazione, secondo l'impostazione offerta da Birger Hjørland, devono elaborare una "Teoria del soggetto" che permetta di produrre rappresentazioni corrette della conoscenza espressa nei libri, anche prendendo in considerazione l'ampio orizzonte costituito dai diversi campi scientifici».

realizzazione di pagine dinamiche per ogni entità di interesse del catalogo, ad esempio un'opera, un soggetto, un autore. In tal modo viene ampliata la possibilità di recupero di un'entità nel suo contesto rappresentando la sua collocazione nell'universo bibliografico del catalogo e usufruendo di collegamenti con gli altri dati nel Web. Un esempio di come potrebbe apparire la pagina di un soggetto è quella fornita dal progetto della Bibliothèque nationale de France [18] (Wenz): a partire da un determinato soggetto viene creata dinamicamente una pagina che offre collegamenti a documenti, film, immagini, risorse multimediali e collegamenti esterni. Visualizzazioni avanzate create a partire dalle entità di interesse del catalogo consentono il recupero dell'informazione bibliografica in un contesto arricchito da altre risorse autorevoli provenienti dal Web. Si passa così da un recupero ispirato all'*information retrieval* ad un recupero in base al modello entità-relazione, che si rivela particolarmente efficace nel raggruppamento dei concetti nella ricerca semantica.

5 Le nuove basi dell'accesso bibliografico

Ciò che, infine, segnerà profondamente il processo di evoluzione del catalogo elettronico e il rapporto utente-catalogo sarà lo spostamento a livello del Web dei meccanismi del controllo bibliografico e dell'organizzazione della conoscenza, ovvero delle principali vie d'accesso al catalogo. I linked data aprono nuove prospettive di accesso all'informazione bibliografica allo scopo di garantirne la fruibilità e il riutilizzo gettando nuova luce sui temi dell'interoperabilità, della pluralità e della convergenza che da decenni sono oggetto di riflessione e che sono alla base di un'adeguata progettazione di ambienti digitali.

[18] http://data.bnf.fr.

A questo punto l'OPAC si trasformerà da strumento autoreferenziale a strumento profondamente integrato nel web e in altri ambienti digitali, realizzando quelle condizioni di interoperabilità da sempre auspicate favorendo così la ricerca e l'uso dei dati bibliografici da chiunque cerchi informazioni nel web e non più dai soli utenti del catalogo (Baker).

Le tecnologie dei linked data potranno facilitare il raggiungimento delle quattro principali tipologie di interoperabilità che consentono la costruzione di ambienti digitali integrati (Leombroni) e cioè quelle dell'interoperabilità tecnologica, semantica, organizzativa e delle risorse umane:

1. *Interoperabilità tecnica*: è la garanzia dell'interoperabilità fra sistemi e applicazioni; essa è resa possibile dall'adozione di una tecnologia unificante come quella fornita dal data model RDF comune a tutto il web;

2. *Interoperabilità semantica*: la tecnologia dei linked data consente alle varie comunità coinvolte di definire concetti il cui significato è condiviso tramite il meccanismo degli URI che collega ai concetti i termini con i quali le diverse comunità li identificano e utilizzano per la ricerca;

3. *Interoperabilità delle risorse umane*: i linked data permettono una convergenza dei dati e dei servizi che attorno ad essi si sviluppano; a tal fine favoriscono l'interoperabilità delle competenze distribuite nel Web Semantico che vengono collegate insieme per lo sviluppo di nuovi servizi basati sui dati;

4. *Interoperabilità organizzativa*: i linked data consentono di sviluppare una convergenza dei servizi delle istituzioni che operano nell'ambito del cultural heritage, della ricerca scientifica, dei servizi culturali, che hanno il delicato compito di cooperare

con le altre comunità del Web affinché venga garantita la qualità e l'integrità e dei dati immessi attraverso idonee procedure di certificazione della provenienza dei dati (Crupi; Morando). Quest'ultima tipologia comprende anche una nuova interoperabilità giuridica relativa alle licenze con le quali i dati vengono pubblicati nel *Web Semantico*, raggiungibile attraverso l'armonizzazione delle licenze d'uso presenti a livello internazionale tramite l'adozione di licenze di utilizzo dei dati più appropriate, e possibilmente, aperte.

Riferimenti bibliografici

Baker, Thomas. «Designing data for the open world of the Web». *JLIS.it* 4.1. DOI: 10.4403/jlis.it-6308. (2013): 63–66. (Cit. a p. 97).

Bateson, Gregory. *Mind and Nature: A Necessary Unity*. New York: Dutton, 1984. (Cit. a p. 77).

Berners-Lee, Tim, James Hendle e Ora Lassila. «» (). (Cit. a p. 90).

Biagetti, Maria Teresa. «Indici semantici e ricerca scientifica». *L'organizzazione del sapere. Studi in onore di Alfredo Serrai*. A cura di Maria Teresa Biagetti. Milano: Sylvestre Bonnard, 2004. 15–29. (Cit. alle pp. 89, 94, 95).

Bianchini, Carlo. «Dagli OPAC ai library linked data: come cambiano le risposte ai bisogni degli utenti». *AIB Studi* 52.3. DOI: 10.2426/aibstudi-8597. (2012): 303–323. (Cit. a p. 80).

Borgman, Christine L. «Why are online catalogs still hard to use?» *Journal of the American Society for Information Science* 47.7. (1996): 493–503. (Cit. a p. 83).

Coyle, Karen. «Linked Data: an Evolution». *JLIS.it* 4.1. DOI: 10.4403/jlis.it-5443. (2013): 53–62. (Cit. a p. 78).

Crupi, Gianfranco. «Beyond the Pillars of Hercules: Linked data and Cultural heritage». *JLIS.it* 4.1. DOI: 10.4403/jlis.it-8587. (2013): 25–49. (Cit. a p. 98).

Di Domenico, Giovanni. «Conoscenza, cittadinanza, sviluppo: appunti sulla biblioteca pubblica come servizio sociale». *AIBS Studi* 53.1. (2013): 13–25. <http://aibstudi.aib.it/article/view/8875>. (Cit. a p. 79).

Fattahi, Rahmatollah. *From Information to Knowledge: SuperWorks and the Challenges in the Organization and Representation of the Bibliographic Universe: Lectio magistralis in Biblioteconomia*. Firenze: Casalini Libri, 2010. (Cit. alle pp. 81, 88).

—. *The relevance of cataloguing principles to the online environment: an historical and analytical study: a thesis submitted in fulfillment of the requirements for the degree of doctor of philosophy in the School of information, library and archive studies*. Sidney: University of South Wales, 1997. http://profsite.um.ac.ir/~fattahi/thesis1.htm. (Cit. a p. 81).

Godbold, Natalia. «Beyond information seeking: towards a general model of information behaviour». *Information Research* 11.4. (2006): 269. <http://InformationR.net/ir/11-4/paper269.html>. (Cit. a p. 83).

Guerrini, Mauro e Tiziana Possemato. «Linked data: a new alphabet for the semantic web». *JLIS.it* 4.1. DOI: 10.4403/jlis.it-6305. (2013): 67–90. (Cit. a p. 78).

Hatop, Götz. «Integrating Linked Data into Discovery». *Code4lib Journal* 21. (2013): 67–90. <http://journal.code4lib.org/articles/8526>. (Cit. a p. 85).

Heath, Tom e Christian Bizer. «Linked Data: Evolving the Web into a Global Data Space». *Synthesis Lectures on the Semantic Web: Theory and Technology* 1.1. (2011): 1–136. <http://linkeddatabook.com/editions/1.0>. (Cit. a p. 78).

Iacono, Antonella. «Towards a new model of OPAC. From information to knowledge». *JLIS.it* 4.2. DOI: 10.4403/jlis.it-8903. (2013): 85–107. (Cit. alle pp. 79, 81).

Klein, Maximilian e Kyrios Alex. «VIAFbot and the integration of library data on Wikipedia». *Code4Lib Journal* 2. (2013): 85–107. <http://journal.code4lib.org/articles/8964>. (Cit. a p. 86).

Kroeger, Maximilian. «The road to BIBFRAME: the evolution of the idea of bibliographic transition into a post-MARC Future». *Cataloging & Classification Quarterly* 51.8. (2013): 1–18. (Cit. a p. 80).

Kulthau, Carol Collier. *Seeking meaning: a process approach to library and information services*. Westport: Libraries Unlimited, 2004. (Cit. a p. 83).

Le Boeuf, Patrick. «A Strange Model Named FRBRoo». *Cataloging and Classification Quarterly* 50.5-7. (2012): 422–438. (Cit. a p. 81).

——. «Foreword». *Cataloging and Classification Quarterly* 50.5-7. DOI: 10.1080/01639374.2012.682001. (2012). (Cit. a p. 81).

Leombroni, Claudio. «Appunti per un'ontologia delle biblioteche digitali: considerazioni sulla Biblioteca digitale italiana». *Bollettino AIB* 44.2. (2004): 115–131. <http://bollettino.aib.it/article/view/4902>. (Cit. a p. 97).

Malmsten, Martin. «Cataloguing in the open - the disintegration and distribution of the record». *JLIS.it* 4.1. DOI: 10.4403/jlis.it-5512. (2013): 417–424. (Cit. alle pp. 79, 85).

Marchitelli, Andrea e Giovanna Frigimelica. *OPAC*. Roma: AIB, 2012. (Cit. a p. 79).

Marcum, Deanna. *A Bibliographic Framework for the Digital Age*. 2011. http://www.loc.gov/marc/transition/pdf/bibframework-10312011.pdf. (Cit. a p. 80).

Morando, Federico. «Legal interoperability: making Open Government Data compatible with businesses and communities». *JLIS.it* 4.1. DOI: 10.4403/jlis.it-5461. (2013): 441–452. (Cit. a p. 98).

Serrai, Alfredo. *Del catalogo alfabetico per soggetti: semantica del rapporto indicale*. Roma: Bulzoni, 1979. (Cit. a p. 89).

——. *Indici, logica e linguaggio: problemi di catalogazione semantica*. Roma: Consiglio Nazionale delle Ricerche - Laboratorio di studi sulla ricerca e sulla documentazione scientifica, 1974. (Cit. a p. 89).

Žumer, Maja e Marcia Lei Zeng. «ISKO-UK Biennial Conference, July 4-5, 2011, London». 4-5 luglio 2011. http://www.iskouk.org/conf2011/ppt/zumer.pdf. (Cit. a p. 92).

Ventura, Roberto. *Il senso della biblioteca*. Milano: Editrice bibliografica, 2011. (Cit. a p. 89).

Wenz, Romain. «Linked open data for new library services: the example of data.bnf.fr». *JLIS.it* 4.1. DOI: 10.4403/jlis.it-5509. (2013): 403–415. (Cit. a p. 96).

Wilson, Thomas D. «Models in information behaviour research». *Journal of Documentation* 55.3. (1999): 249–270. <http://informationr.net/tdw/publ/papers/1999JDoc.html>. (Cit. a p. 83).

Yee, Martha M. «Beyond the Opac: Future Directions for Web-Based Catalogues». *Beyond the OPAC : future directions for Web-based catalogues*. 2006. http://www.nla.gov.au/lis/stndrds/grps/acoc/papers2006.html. (Cit. a p. 80).

Yu, Holly, Young Margo. «Models in information behaviour research». *The Impact of Web Search Engines on Subject Searching in OPAC* 23.4. (2004): 168–180. <www.ala.org/ala/lita/litapublications/ital/volume23a/number4a/yu.pdf>. (Cit. a p. 83).

ANTONELLA IACONO, Università di Roma "La Sapienza".
antonella.iacono@fastwebnet.it

Iacono, A. "Dal record al dato. Linked data e ricerca dell'informazione nell'OPAC". *JLIS.it*. Vol. 5, n. 1 (Gennaio/January 2014): Art: #9095. DOI: 10.4403/jlis.it-9095. Web.

ABSTRACT: The application of linked data to catalogue data opens up new perspectives for OPAC development. The new relationships between bibliographic and other data will produce cataloguess that are deeply integrated with the Web, giving a decisive impact on the structure of catalog records, on access and on the functions that the catalog can play in the wider global information space. In this paper the author explores the possibility the new record deconstructed and connected with the other data on the Web is able to facilitate the creation of knowledge in the use of the catalogue. The author then analyzes the potential of application of linked data to the catalogue with regards to the capabilities for research, the new possibilities of semantic search and the ways to access bibliographic data.

KEYWORDS: OPAC; Library linked data; Semantic web; Semantic search; Subject search.

Submission: 2013-10-03
Accettazione: 2013-12-10
Pubblicazione: 2014-01-01